2011 Third International Conference on Multimedia Information Networking and Security

Rumor Propagation in Online Social Networks Like Twitter—A Simulation Study

Dechun Liu
Department of E-Business, Nanjing University
Nanjing, China
daphnezhm@gmail.com

Xi Chen*
Department of E-Business, Nanjing University
Nanjing, China
doctor_chan@163.com

Abstract—**Rumor has drawn much of the attention of researchers, considering its importance and influence, as well as its complexity. Under different circumstances, the process and results of rumor spreading are also different. The method of agent-based modeling is considered a good way to look into the process of rumor propagation. In this study, we put the theme under the circumstances of online social networks like twitter, which is different from traditional social network websites. Assuming that the structure of these websites is a kind of scale free network, we build a specific agent-based model using NetLogo based on the propagation model of SIR. Taking consideration of the unique property of twitter-like websites in information flow, max spread time and the mechanism of people believing the rumor, we draw the conclusion that compared to the traditional online social networks, the rumor propagation in twitter-like websites is of more efficiency, unpredictability, and a tendency to start the rumor again.**

Keywords-rumor spreading; agent-based modeling; social network

I. INTRODUCTION

A lot has been studied in rumor spreading field, some of which build numeric models to describe the spreading procedure. Those models are mostly simulating the rumor spreading through word-of-mouth in real society.[1] As a matter of fact, the rumor spreads locally in a region and spills over out of the region where it has never been spread with movement of people and information.[2] This is a classic procedure of rumor spreading in real world. With the Internet popularized, Internet surfers have more access to the news. Besides BBS, there are blogs and SNS(social network sites) where users can broadcast information to their friends and others who care about it.[3] Nowadays, a new form of web2.0 website has become popular. It's called microblog, and the most famous microblog site is Twitter.com in USA and the world, the registered users of which is up to 175million. In China, people are trying to catch up with world's pace, with more and more people using microblogs, among which there are t.sina.com and t.qq.com and so on. Compared to other social network sites like the most famous site named Facebook.com, and Renren.com in China, people use microblog mostly to broadcast news and their status within 140 words.[4] Another characteristic that distinguishes the microblog from SNS is that there are two kinds of relationship between users instead of one—follow each other, or follow in one direction. This characteristic of microblog makes it different when transmitting information through users. In the following paragraphs, I will first describe the Twitter-like user network, how it forms and why it is a scale-free network. Secondly I will introduce several classic models that demonstrate the spreading of rumor, and how it will change when it meets the Internet, especially how it works in Twitter-like internet network. Thirdly I will come up with several assumptions and use them to build a spreading model. At last, I will test the model in software named NetLogo, to see what result the model brings, and what useful suggestion it will give us.

Compared to former studies, I make some differences to extend their models and research. Though there are studies concerning rumor spreading in scale-free network, they just consider the user structure an undirected graph. In my study, I define a follow-back-rate to describe the true situation in Twitter-like microblog websites. In this kind of websites not all users hear from each other, thus the rumor spreading will result in different ways.

What's more, classic models don't include the mechanism for people to determine whether or not to believe the rumor. In real world however, it's critical to find out the ratio of people who believe the rumor, which is obviously related to the depth of rumor influence.

II. LITERATURE REVIEW

A. The Twitter-like website network

Like other social network websites, Twitter is a website where users register, fill in their personal profiles, choose some of the other users who are already in the website to follow, and then broadcast their thoughts or moods in a limit of 140 words. When the users you are following say something new, they will appear in your timeline at the homepage of Twitter.[4] But those words published by users you are not following won't appear in your timeline, in other words, you won't be able to find them since there is a large number of users in Twitter. When you see the words published by others, you can retweet it so that people following you can see it in their timelines[5], which is one of the best characteristics that twitter has.

But the difference between twitter and Facebook lies in the fact that when you want to follow someone named A in Facebook, you must send a quest and wait for A's confirmation. If not, you won't be able to track him/her. But when user A agrees to your request and you two are friends now, he/she can see what you say in his/her timeline,

* Corresponding author: Xi Chen, E-commerce department, Nanjing University (E-mail:doctor_chan@163.com)

978-0-7695-4559-2/11 $26.00 © 2011 IEEE
DOI 10.1109/MINES.2011.109

regardless of whether or not he is willing to. However, within Twitter you can feel free to follow someone named B regardless of his/her permission and B will not receive what you say in his/her timeline until he/she chooses to follow you back. In other words, the relationship between users in Twitter-like microblog websites is directed, therefore information is transmitted in a directed line.[6]

Apart from what mentioned above, the shaping of Twitter structure is more or less the same as Facebook. In one's profile page, others can see who is following him/her and who he/she is following. When you decide to follow someone, you may consider how many people are following him/her. The more followers he/she has, the more likely you may choose to follow him/her, because there must be some reasons for them to make the same decision. As a result, there should be a few people who have more followers above the average. Therefore it's reasonable to assume that these people have more influence on others when they spread some information or broadcast their opinions.

The network described above is called scale-free network in the theory of social network. Albert etc. first brought it up and found that in real society, the topology of the network is not ER or random network, but a scale-free network, where preferential attachment was introduced,[7] which, in common words, is that the probability A connects to B is depended on how many connections B already has. They also found that since world-wide web forms a large directed graph, the topology of this graph has a number of universal scale-free characteristics. Java, A., X. D. Song, et al. looked into the user structure network and found that Twitter social network has a power law exponent of about -2.4 which is similar to the Web and blogosphere.[4] In other words, the topology of user structure in Twitter-like microblog website is likely to be seen as a scale-free network. Actually, Tong yang Yu collected 4036 users of a SNS website and their relationships to form an undirected graph.[8] He studied some basic attribute of the graph, including the degree distribution, clustering, and network core, finding that the network of SNS users is likely to be a scale-free network. Though Twitter-like website forms a directed graph, the main attribute is the same.

B. Rumor spreading models

Many rumor spreading models are built on the model of disease spreading. Zanette use the theory of complicate network to study the rumor spreading, building a spread model in a small-world type of network.[9] Moreno etc. extended the model by building it in a scale-free network.[10] They divided people into three different groups. One is Igorants, who have never heard of the rumor. Second is Spreaders, who spread the rumor. And third is Stiflers, who know about the rumor but don't spread it. The rumor spreads through spreaders and igorants, by every step one spreader spreads the rumor to one or some of his neighbors. If the receiver is an igorant, he will turn to be a spreader at the possibility of λ. If the receiver is a stifler or also a spreader, the resource spreader will turn to be a stifler at the possibility of $1/\alpha$. Here α means the number of neighbor spreader or stifler one spreader may spread the rumor to. This model makes rumor spreading in a scale-free social network descriptive, and draws some useful conclusions.

But this model is used in real society, where people can walk around and tell what he knows to his acquaintances, and the spreader knows who the receiver is and what attitude the receiver holds to the information at the same time he spreads it, then the spreader can make decisions whether to spread the information again or not. And in the real world, one spreader can keep spreading the rumor to as many people as he wants as he moves around, until he face up to enough amount of spreaders or stiflers, that's when he stops spreading. But for users of a website, especially a website like Twitter or other social network sites, things are different. In this kind of websites, users only care about what is discussed among who they follow, and only the ones he/she follows show new messages in his/her timeline. Since information flow between users is a one-way type, the spreader can't know what the followers think when he/she broadcasts the information. So the spreading itself can't be the reason why a spreader changes his mind. Another thing we must pay attention to is that a user can follow as many other users as he/she feels like to, so there will be many messages displayed in one's timeline. There are possibilities that one may miss some of them.

In social network websites, users usually don't know each other as well as in the real world, so there must be new rules to help them decide whether to believe the rumor or not. Some studies use game theory.[11] Some use decision list.[2] Considering the unique characteristic the Twitter-like websites own, I will use a mechanism called Influence, which aroused by former study. Xia and Huang introduced CIF and Behavior Update Rule, by which people will believe the rumor if absolute value of cumulative influencing force (CIF) of him/her about the rumor is equal or bigger than the threshold of him/her.[11] As in a scale-free network, there are always some users who have followers above the average. One of the reasons maybe lie in this: he/she is a public person and may have many public resources, so that people are used to listening to what he/she says and discuss it. As a result, these people have more influence on others than those who have fewer followers.

According to the above, I bring up my assumptions and build a model based on a Twitter-like website user structure.

III. MODEL

A. The assumptions:

- Twitter-like websites have a user structure of a scale-free network.
- The network is a directed network, in which users always searching for those who have more influence, in other words, have more users followed, to follow. In order to gain more attention and more influence, those users who have many followers usually follow back. So there are two relationships in this network: friendship, in which two users follow each other; and followership, in which one user follows the other, but the other does not follow back.

- The network structure will never change once it forms. Although it's dynamic in real world, users usually change who they are following since there are new users adding in. As the rumor spreads in a fast speed, a little change in network structure won't change the final state of the rumor spreading, so I decide not to take it into account.
- Users have fully access to their followers and those they follow. That means one can get the information about how many followers they have and how many users they follow.
- The more follower one user has, the more influence he/she will have on his/her followers.

B. The model:

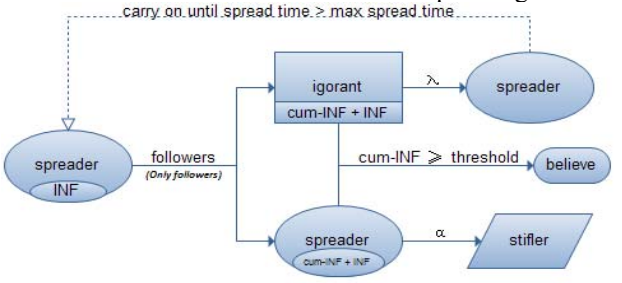
The Twitter-like website user structure is a scale-free network. I use a directed graph to describe it. In the graph, if user A follows user B, then information comes from B will flow to A. Then there is a directed link from B to A. I define the out-degree of user A as the number of followers A has, and the in-degree as the number of users that A follows.

There are also three groups of people. One is spreaders, who spread the rumor through the use of microblog. One is stiflers, who won't spread the rumor any more. The last is igorants, who have never heard of the rumor. I want to discover the result of rumor spreading in Twitter-like website, and the result dose not only mean how many igorants or stiflers at last, I also want to know how many of them believe the rumor, how many do not believe but still spread the rumor, and how many of them do believe but don't spread it. Since the structure of the network is different from real world and there are advantages if we make rules for users to decide whether to believe the rumor, to keep spreading or to stop spreading, we can easily make some rules in order to see the results.

The rumor starts where a few of the users become spreaders. Then they tweet about the rumor, so their followers can see it. The fact is that the spreaders don't have the access to know what group their followers belong to, so they can't decide what step to take afterwards. Considering this, it makes sense if we let the receiver decide what to do, because he knows exactly which group both he and the spreader belong to. If the receiver is an igorant, he will become a spreader at the chance of λ. If the receiver is a spreader, he will become a stifler at the chance of α. In this case, if the spreader never hear from others again, he is supposed to spread the rumor ever since without a stop. So each spreader has a max spreading times. The number describes one's passion to spread the rumor.

Then it comes to the believing procedure. I define an attribute for each user to represent the influence of them. I use INF to represent it, by calculating out-degree divided by in-degree. The more followers one has, the bigger his INF will be. When user A heard a rumor from different users, he/she will feel the INFs and sum them together. He will then compare the accumulative result to a threshold, which will help him decide whether to believe the rumor or not. Here I set the threshold between 0 and 10 times of his/her own INF. In common sense, one tends to judge others' influence by his/her own, thus making the INF a relative concept. So the threshold is a variable that related to one's INF. I just define it between 0 and one's 10 times INF. Other ways to define it may change the result.

In the end of the spreading, there will be no more spreaders, only stiflers and igorants. Both of the groups contain people that believe the rumor and people that don't. What I care about is the factors that can affect the proportion of people who believe the rumor. Considering the dynamic characteristic of microblog website, the structure of it changes quite often, there are always new users joining in and new links created. Those messages will appear in new users' timelines and it will restart the rumor spreading.



INF = out-degree/in-degree. threshold = 0 ~ 10×INF. max spread time is pre-set

C. The process to build the model:

First steps to build a BA Scale-free Directed Network

- starts with a few agents(m0 agents), connect each of them to one of the others
- at every time step, add a new agent to the network, and make m(m⩽m0) links from the agents already exist. The adding rule is: the probability that the new agent will be linked from agent i depends on the out-connectivity out-ki of that agent, such that
  $\prod out - k_i = out - k_i / \sum_j out - k_j$
- when an old agent is connected to a new agent, there is a possibility (follow back rate-fbr) that it will make a directed link from the new one.
- Calculate each of every agent's INF, by out-degree/in-degree.

Second steps to simulate the spreading of a rumor:

- start the rumor by making a few of the agents become spreaders
- define max-spread-time. Each spreader spread the rumor no more than the max time of time step.
- Set threshold randomly between 0 and INF times the threshold multiplier.
- The rumor spread from the spreader to his out-link-neighbors. Assume A is one of them. If A is an igorant, he will have a chance of p to become a spreader. If A is a spreader already, he will have a chance of r to become a stifler. If A is a stifler, there is no chance that the rumor can reach to him.
- Each time one agent hears about the rumor, he will add the resource agent's INF into his accumulative-INF (abbr. cum-INF). At the end of spreading, if

one's cum-INF is bigger than his believing threshold, he will believe the rumor, otherwise he will not.

The software interface of the model is shown in Fig. 1.

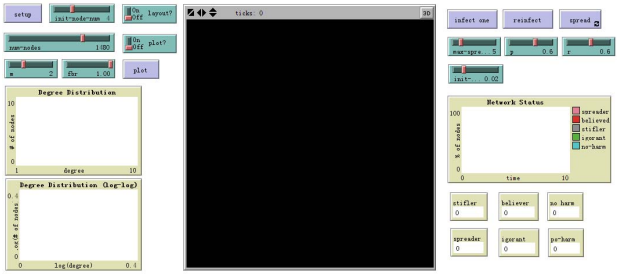


Figure 1. The software interface.

IV. RESULTS AND DISCUSSION

A. Results

After creating the network, I check the Degree Distribution using original data and log data (see Fig. 2 and Fig. 3). It fits well with scale-free network.

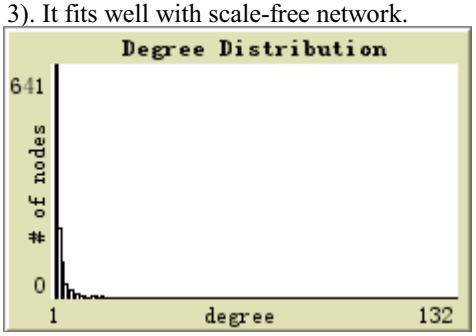


Figure 2. The (out-)degree distribution of one of the results simulated by the model.

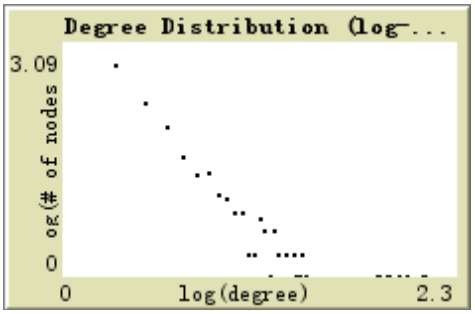


Figure 3. The (log)degree.

1) The influence of the follow back rate

If the follow back rate is zero, then the information flows in one direction. If the follow back rate is 100%, then the network will turn to be a traditional social network site.

From many runs, the rate of agents that believe the rumor is statistically significantly different. In traditional social network sites, where there is only one kind of user relation, the final ratio is fairly low, between 41% and 44%. But in Twitter-like website, where there are two kinds of user relations, the believer ratio is usually high. When the follow back ratio is 0, to the extreme, the believer ratio is between 71% and 76%. The difference is significant in 0.001 confidence level (see Fig. 4 and Fig. 5).

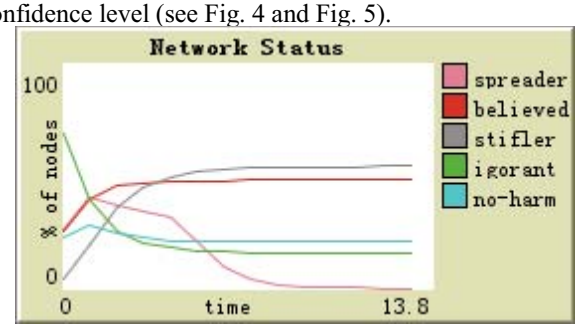


Figure 4. The result displayed when fbr=1.

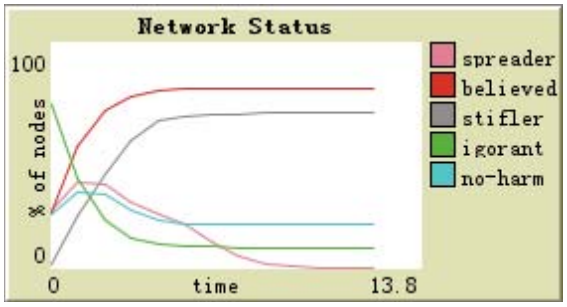


Figure 5. The result displayed wen fbr=0.

2) Other ratios

Those igorants that have never heard of the rumor is important for the rumor spreading. It represents the efficiency of the spreading. When fbr is 1, that's in traditional social network, the ratio of igorants is around 20%. In Twitter-like sites, this ratio falls down to around 10%. That means, in Twitter-like websites, the spread efficiency is higher, thus there are fewer users that don't know about the rumor.

Believing the rumor or not, users can spread the rumor as well. Therefore there are users who don't believe the rumor but still spreading it, and also users who do believe the rumor but just don't feel like spreading it. In the end, if there are users who are not stiflers and at the same time believe the rumor, he will continue to spread the rumor if there are new users joining in the network. This ratio, in both networks, is between 15% and 20%, when p is more than 0.5 even if r is very big. But when r is very small, that means there is little chance that a receiver becomes a stifler, the ratio of potential spreader is very big, reach 50%.

Even if not every variable is extremely reaching the limit value of its intervals, there are chances that Fig. 6 appears, in which there are big number of users who have not heard of the rumor. The reason may be that the rumor starter have little followers and his followers have little followers either. This represents the situation that the rumor is probably faked and only a small field of people care about it. It won't happen in undirected network because the rumor spreads in both directions.

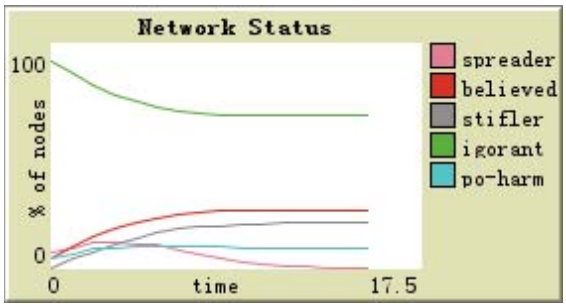


Figure 6.

B. *Contribution to research and practice*

In this study I extend the classic rumor spreading model by building a new model on the base of scale-free network taking two kinds of links between people into consideration. A mechanism of believing is also brought into the model, thus transferring the attention for the ratio of stiflers into that of believers. In classic model, the final ratio of the three groups of people is independent with the degree of rumor source. But in this two-kind-of-relationship network, there are chances that the rumor spreads in a small area of people and the ratio of three groups of people is different. That is caused by the one-directed link, where the rumor spreads only in one direction.

As for microblog websites, they can't control the follow back rate, and they probably don't have the intention to do so. It is normal to see that there are many public figures being active users in microblog websites, who have a large number of followers yet follow little. This fact will easily cause explodes of rumor and make people believe the rumor. One way to change this is to increase follow back rate, let people listen more from others, therefore lower their influences. Only by doing so can a microblog website become "a channel to express", instead of "a channel to hear".

Besides, the potentially harmful people will carry on the damage caused by the rumor. This kind of people will always exist, no matter how the variables will change. But if r is big, there are a relatively small number of potentially harmful people. So it is important for the users in microblog websites to improve their ability to tell right from wrong and stop following blindly.

V. LIMITS AND FUTURE WORK

During the simulation, I found that there would always be spreaders if I didn't use the max-spread-time to control it. Though users do have a limited passion to spread a rumor, but setting a threshold is not what I want and may change the result a little bit.

The threshold that a user depends on to decide whether to believe the rumor or not is not set by enough literature. This threshold is important for most ratios that I mentioned. I simply set it randomly between 0 and one user's INF times a variable named threshold multiplier. Reasonably, the threshold is somehow related to a user's own INF, and the larger the INF is, the larger the threshold will be. The probability that a public person is going to believe what an ordinary person says is obvious lower than the other way around. As it may be, there are also unpredictable factors. That's why the random method is used. But this is only the assumption. A method to test it is extracting data from a microblog website, examining the relation between the influences of a user and his tendency to believe someone else.

The mechanism of believing is not the only thing that has limits. The definition of INF is a rough estimate for the real influence of a user. It should be including the ability to broadcast information, the activeness and the coverage of the user, etc.

This simulation is based on the network that created using programming language. In the future work, the spreading procedure can be simulated on a real network that generated by a real Twitter-like website. The results can be tested then to see if they are the truth.

ACKNOWLEDGEMENTS

The work was supported in part by the National Natural Science Foundation of China under Grant No. 70901039 and 70732002, National Postdoctoral Science Foundation of China under Grant No. 20090450144 and National Postdoctoral Science Special Foundation of China under Grant No.201003565, Jiangsu Postdoctoral Science Foundation under Grant No. 0901104C.

REFERENCE

[1] Fang, Z., S. Guang-ya, and L. Pi, A Survey for Rumor Propagation Models. Complex Systems and Complexity Science, 2009. 6(4): p. 1-11.

[2] Kawachi, K., et al., A rumor transmission model with various contact interactions. Journal of theoretical biology, 2008. 253(1): p. 55-60.

[3] Nai-peng, C. and H. Xian, A Research of Rumor Propagation in Internet Communication. Information Studies: Theory and Application, 2004. 27(6): p. 586-575.

[4] Java, A., et al., Why We Twitter: An Analysis of a Microblogging Community. Advances in Web Mining and Web Usage Analysis, 2009. 5439: p. 118-138.

[5] Boyd, D., S. Golder, and G. Lotan, Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. 43rd Hawaii International Conference on Systems Sciences Vols 1-5 (Hicss 2010), 2010: p. 1657-1666.

[6] Suh, B., et al. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. 2010. IEEE.

[7] Barabási, A.L., R. Albert, and H. Jeong, Mean-field theory for scale-free random networks. Physica A: Statistical Mechanics and its Applications, 1999. 272(1-2): p. 173-187.

[8] Tong-yang, Y., Multi-agent Simulation Approach to Information Diffusion Among Network, 2010, Huazhong University of Science and Technology: Wuhan.

[9] Zanette, D.H., Dynamics of rumor propagation on small-world networks. Physical Review E, 2002. 65(4): p. 041908.

[10] Moreno, Y., M. Nekovee, and A.F. Pacheco, Dynamics of rumor spreading in complex networks. Physical Review E, 2004. 69(6): p. 066130.

[11] Xia, Z.Y. and L.L. Huang, Emergence of Social Rumor: Modeling, Analysis, and Simulations. Computational Science–ICCS 2007, 2007: p. 90-97.